Proc. Natl. Acad. Sci. USA
Vol. 78, No. 1, pp. 205–209, January 1981
Biochemistry

Template recognition sequence for RNA primer synthesis by gene *4* protein of bacteriophage T7*

(DNA initiation/primase/lagging-strand synthesis/Okazaki fragments/DNA synthesis)

STANLEY TABOR AND CHARLES C. RICHARDSON

Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115

Communicated by Eugene P. Kennedy, October 6, 1980

ABSTRACT **The gene *4* protein of bacteriophage T7 recognizes specific sequences on single-stranded DNA and then catalyzes the synthesis of tetraribonucleotide primers complementary to the template. With φX174 DNA as a template, the gene *4* protein enables T7 DNA polymerase (deoxynucleosidetriphosphate:DNA deoxynucleotidyltransferase, EC 2.7.7.7) to initiate DNA synthesis at 13 major sites. DNA sequence analysis of the 5′ termini of the newly synthesized DNA shows the predominant recognition sequences for the gene *4* protein to be 3′-C-T-G-G-G-5′ or 3′-C-T-G-G-T-5′; the products of synthesis at these sites are RNA primers having the sequences pppA-C-C-C or pppA-C-C-A. The gene *4* protein can also synthesize primers at the sequences 3′-C-T-G-G-$\overset{A}{C}$-5′ and 3′-C-T-G-T-N-5′, although these sites are used less than 10% as frequently as the predominant sites. Comparison of the utilization of primer sites suggests that the gene *4* protein binds randomly to single-stranded DNA and then translocates along the DNA in a unidirectional 5′-to-3′ direction with regard to the DNA strand in search of recognition sequences. Models are presented for the role of the gene *4* protein in the initiation of lagging-strand synthesis and in the initiation of DNA replication at the origin.**

Recent studies *in vitro* have provided detailed information on the enzymatic mechanisms involved in DNA replication. In this paper, we use bacteriophage T7 as a model system (for review, see ref. 1) to describe the mechanism by which two phage proteins, T7 DNA polymerase (deoxynucleosidetriphosphate:DNA deoxynucleotidyltransferase, EC 2.7.7.7) and gene *4* protein, synthesize Okazaki fragments during T7 DNA replication. The T7 DNA polymerase is composed of two subunits (2–4): the phage gene *5* protein (M_r 87,000) and *Escherichia coli* thioredoxin (M_r 12,000). The gene *4* protein, a single polypeptide, is found in two forms with molecular weights of 58,000 and 66,000 (5, 6).

The gene *4* protein has a dual role at the replication fork. First, it enables T7 DNA polymerase to use duplex DNAs as templates (5, 7, 8). T7 DNA polymerase catalyzes the polymerization of dNTPs, while the gene *4* protein catalyzes the hydrolysis of NTPs to NDPs and P_i (9), a reaction that is coupled to DNA synthesis. Presumably the energy derived from hydrolysis is used in unwinding the DNA helix ahead of the replication fork, because inhibition of the hydrolysis reaction results in a cessation of DNA synthesis on duplex DNA. The gene *4* protein initiates lagging-strand DNA synthesis by catalyzing the synthesis of RNA primers (6, 10–12). T7 DNA polymerase, like all DNA polymerases, is unable to initiate DNA synthesis *de novo*. However, it readily uses the RNA primers synthesized by the gene *4* protein to begin synthesis. Lagging-strand synthesis is the result of multiple priming events, with the transient production of Okazaki fragments.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

The RNA primers synthesized by the gene *4* protein *in vitro* are predominantly tetraribonucleotides having the sequences pppA-C-C-A or pppA-C-C-C (6, 12). By what mechanism does the gene *4* protein synthesize these tetraribonucleotide primers? A possible mechanism might be analogous to the template-independent tRNA nucleotidyltransferase that specifically adds the -C-C-A terminus to tRNAs (13). However, the fact that single-stranded DNA is required for primer synthesis suggests a template-mediated reaction (10, 11). Scherzinger *et al.* (10) have inferred site-specific primer synthesis on the basis of a restriction enzyme analysis of the product synthesized on single-stranded circular DNA. In this communication we map the precise locations of primer initiation sites to show that there is a distinct template recognition sequence that the gene *4* protein must recognize before it synthesizes a RNA primer.

MATERIALS AND METHODS

DNA and Nucleotides. Bacteriophage φX174 am3 DNA was isolated by the method of Hutchison and Sinsheimer (14). φX174 replicative form DNA was obtained from New England BioLabs. [γ-^{32}P]ATP and cytidine 3′-phosphate 5′-[^{32}P]phosphate ([^{32}P]pCp) were from New England Nuclear.

Enzymes. Gene *4* protein was fraction V (50% pure) of Kolodner *et al.* (5). T7 DNA polymerase (80% pure) was prepared from $T7_{3,4,6}$-infected *E. coli* D110 as described by Adler and Modrich (15). Bacteriophage T4 polynucleotide kinase and *E. coli* alkaline phosphatase have been described (16). Restriction enzymes were obtained from New England BioLabs.

RNA-Primed DNA Synthesis. The reaction mixture (0.1 ml) contained 40 mM Tris·HCl (pH 7.5), 10 mM $MgCl_2$, 50 mM dithiothreitol, 0.3 mM each dNTP, 0.12 mM each rNTP, including [γ-^{32}P]ATP (50 Ci/mmol; 1 Ci = 3.7×10^{10} becquerels), 6 nmol φX174 DNA, and 0.02 ml of a solution containing 2 units of T7 DNA polymerase and 2 units of gene *4* protein in 10 mM Tris·HCl (pH 7.5), 10 mM 2-mercaptoethanol, and bovine serum albumin at 0.5 mg/ml. After incubation at 30°C for 10 min, 0.01 ml of 0.5 M EDTA and 0.1 ml of redistilled neutralized phenol were added at 0°C. Radioactive ATP was separated from the product DNA by gel filtration through a 1.0-ml Sephadex G-100 column in 20 mM Tris·HCl (pH 7.8)/2 mM EDTA/100 mM NaCl, and the DNA in the void volume was precipitated with ethanol. The restriction enzyme digestion reaction mixture (0.1 ml) contained 10 mM Tris·HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 10 mM NaCl, bovine serum albumin at 0.1 mg/ml, the end-labeled φX174 DNA, and 20 units of *Hae* III. Incubation was at 37°C for 60 min. The DNA

* This is paper no. 19 in a series entitled "Replication of Bacteriophage T7 Deoxyribonucleic Acid." The previous paper is ref. 27.

was extracted with phenol, precipitated with ethanol, and applied to a 6% polyacrylamide gel containing 7 M urea.

End-Group Labeling of DNA Product. The product DNA bearing 5′-terminal RNA primers was incubated with 0.1 M KOH at 37°C for 20 hr. The reaction mixture was neutralized and the DNA was precipitated with ethanol. The kinase reaction mixture (25 μl) contained 50 mM Tris·HCl (pH 8.0), 5 mM dithiothreitol, 0.1 mM spermidine, 10 mM $MgCl_2$, 4 μM [γ-^{32}P]ATP (3000 Ci/mmol), the alkali-treated DNA, and 10 units of T4 polynucleotide kinase. Incubation was for 30 min at 0°C, in order to limit the kinase-catalyzed exchange reaction between the γ phosphate of ATP and the 5′ phosphate of nicked DNA (17). Radioactive ATP was removed by Sephadex G-100 gel filtration. 5′-Terminal-labeled fragments were annealed with the φX174 template by incubation in 10 mM Tris·HCl (pH 7.5)/1 mM EDTA/50 mM NaCl at 67°C for 30 min.

RESULTS

Scheme for the Determination of Gene *4* Protein Recognition Sites. Our objectives were to determine (*i*) whether priming was occurring at specific sites and (*ii*) whether such sites shared sequence homology. The principle of the method is to map, at the nucleotide level, the positions of initiation events that occur in the presence of the gene *4* protein and T7 DNA polymerase on single-stranded circular φX174 DNA, a DNA molecule whose nucleotide sequence is known (18). The scheme for the determination of gene *4* protein recognition sites is shown in Fig. 1. Using φX174 DNA as a template for T7 DNA polymerase, DNA synthesis is dependent upon RNA primer synthesis catalyzed by the gene *4* protein (6, 10, 11). DNA synthesis proceeds around the circular template, giving rise to a population of double-stranded circular molecules. Because all the primers contain ATP at their 5′ ends (6, 11), the newly synthesized molecules can be uniquely labeled at their 5′ ends with [γ-^{32}P]ATP. Provided initiation occurs at specific sites, discrete size classes of labeled fragments will be produced by treatment with a restriction endonuclease; each initiation site results in a labeled fragment whose 5′-labeled end corresponds to the site of initiation. Thus the number of initiation sites will correspond to the number of radioactively labeled bands on a denaturing gel.

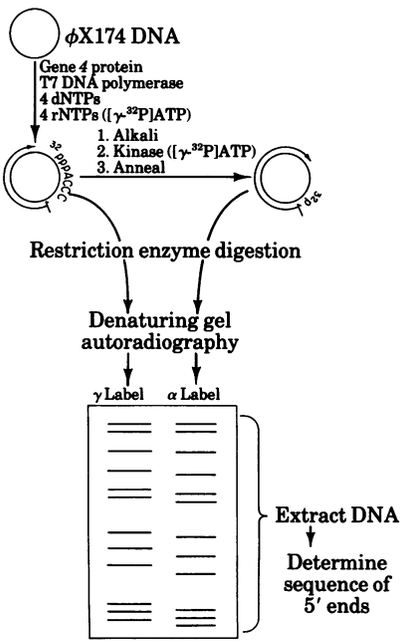


FIG. 1. Scheme for the determination of gene *4* protein recognition sites.

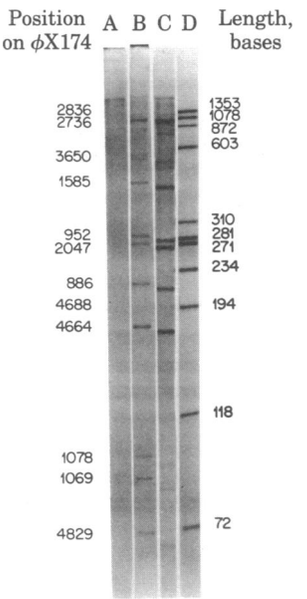


FIG. 2. Site-specific initiation on φX174 DNA. Samples were run on a 6% polyacrylamide gel containing 7 M urea (19). Lane A contains undigested newly synthesized linear φX174 molecules labeled with [γ-^{32}P]ATP at the 5′ ends of their RNA primers. In lane B, the [γ-^{32}P]ATP-labeled molecules have been digested with *Hae* III. In lane C the molecules have been treated with alkali to remove the RNA primers, followed by kinase and [γ-^{32}P]ATP to label the α position of the junction deoxyribonucleotides, prior to digestion with *Hae* III. Lane D contains molecular weight markers: φX174 replicative form I was digested with *Hae* III, and the 5′ termini were labeled with ^{32}P (19). The number of nucleotides in each of the standards is shown on the right. The position on the φX174 template of the initiation site for each fragment digested with *Hae* III, as determined in the subsequent figures, is shown for reference on the left.

Alternatively, the 5′ ends of the DNA portion of the newly synthesized molecules can be radioactively labeled by first removing the RNA primers with alkali and then labeling the 5′-terminal hydroxyl groups with [γ-^{32}P]ATP and T4 polynucleotide kinase. The labeled fragments must be annealed with the φX174 DNA prior to incubation with the restriction enzyme. Because the RNA primers are predominantly four nucleotides in length (6, 12), this procedure yields labeled molecules that are reduced in size by four nucleotides. A major advantage of labeling the 5′ termini of the DNA itself is that one can directly determine the sequences of the 5′ ends; the sequences of molecules that bear 5′-end-labeled RNA primers cannot be determined by standard Maxam–Gilbert techniques (19) because the RNA primers are completely degraded by the alkali used in this procedure. In order to definitively determine the precise nucleotide where each initiation event occurs, it is necessary to determine the sequence of the 5′ terminus of the labeled fragment. The DNA in each band observed by autoradiography is therefore extracted and its nucleotide sequence is determined.

Thirteen Predominant Sites on φX174 DNA Are Recognized by the Gene *4* Protein. φX174 was used as a template in a reaction mixture containing the gene *4* protein, T7 DNA polymerase, and the four dNTPs and rNTPs ([γ-^{32}P]ATP). DNA synthesis on the φX174 molecules led to the formation of duplex circular molecules; all of the 5′-terminal-labeled product migrated on a denaturing gel at a position corresponding to full-length linear φX174 DNA (Fig. 2, lane A). When these molecules were digested with the restriction enzyme *Hae* III (11 cuts on φX174), 13 predominant bands were observed (Fig. 2, lane B). Other faint bands were present, which correspond to minor gene *4* protein recognition sites (see *Discussion*). This result strongly suggested that the gene *4* protein recognized some ele-

ment that occurred 13 times in the ϕX174 sequence. To determine whether this represented a specific sequence common to these sites, the precise location on the template where each initiation event occurred was identified by determining the sequence of the 5′ terminus of each of the bands.

DNA Sequence at 5′ End of a Gene *4* Primer Initiation Site. In order to determine the sequence at the 5′ ends of the synthesized fragments, we removed the γ-^{32}P-labeled primer and replaced it with an α-^{32}P label. After treatment of the product with alkali the 5′ terminus of the DNA was radioactively labeled with polynucleotide kinase and [γ-^{32}P]ATP. The resulting banding pattern (Fig. 2, lane C) is essentially that obtained when the newly synthesized DNA bears ^{32}P-labeled RNA primers (Fig. 2, lane B), except for the faster mobility of the former due to the absence of the tetraribonucleotide primer.

Each of the α-^{32}P-labeled fragments shown in Fig. 2, lane C, was extracted, and its nucleotide sequence was partially determined. An example of the sequence from a single site is shown in Fig. 3. When the sequence complementary to the 5′ end of the fragment is identified on the ϕX174 template (position 4664 in Fig. 3), it is seen that immediately adjacent to the first nucleotide present on the sequencing gel (i.e., equivalent to nucleotides −1, −2, −3, and −4) is found the sequence 3′-T-G-G-T-5′, the sequence complementary to a pppA-C-C-A primer. Because the tetranucleotide primer has been removed with alkali, the first nucleotide detected on the sequencing gel corresponds to the deoxyribonucleotide at the RNA-DNA junction.

An unambiguous sequencing pattern is dependent upon a unique end for all 5′-labeled junction deoxyribonucleotides at a given primer site. Thus the degree of homogeneity in the RNA primer length is reflected by the homogeneity of the sequencing gel. For the initiation site shown in Fig. 3 (position 4664), the sequencing gel is unambiguous, so the RNA primers must be homogeneous in length.

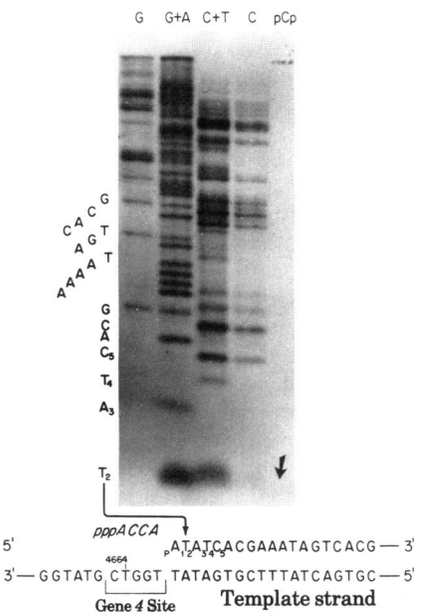


FIG. 3. Sequence at the 5′ end of the fragment initiating at site 4664 on ϕX174 DNA. The α-^{32}P-labeled fragment shown in Fig. 2, lane C, that corresponds to 171 bases in length was extracted from the gel by using a modified Instrumentation Specialties Company electroelution procedure, and its sequence was determined by Maxam–Gilbert techniques (19). The far right lane contains the marker [^{32}P]pCp, which migrates at the same position as the second nucleotide from the 5′ end of the fragment. The tetraribonucleotide primer, which has been alkali degraded in a previous step, is shown in italics. The gene *4* protein recognition sequence (shown in brackets) is based on the homologous sequence in all 13 predominant initiation sites on ϕX174 DNA, as shown in Fig. 4.

Position	Sequence (3′ 5′)	*Hae* III fragment, bases
886	CTGCTGGTCCCG	216
952	CAACTGGTGGAT	282
1069	AGACTGGTCGTT	89
1078	CTGCTGGTTAGA	98
1585	CGTCTGGGTATT	411
1789	GTGCTGGTCTTT	12
2047	TGACTGGGAGTC	270
2736	CCGCTGGTAAGT	959
2836	AATCTGGTTTGG	1059
3650	TGACTGGTCGGC	520
4664	ATGCTGGTTATA	175
4688	ACGCTGGGAGCC	199
4829	CTGCTGGTTTTA	69

Predominant recognition sequence 3′—CTGGG—5′ (T)
RNA primer pppACCC—dN→ (A)

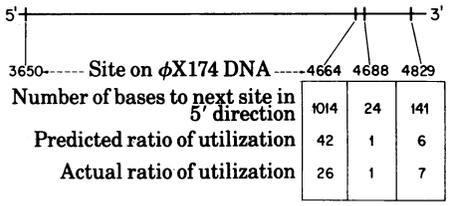
FIG. 4. Predominant gene *4* protein priming sites on ϕX174 DNA. Shown are the nucleotide sequences of the template ϕX174 DNA in the region of each of the 13 gene *4* protein initiation sites. The actual sequences analyzed were the 13 unique 5′ ends of newly synthesized DNA complementary to the sequences shown. The initiation fragments used to obtain the sequences were those shown in Fig. 2, lane C, with the exception of the fragment from initiation site 1789, which, due to its small size after *Hae* III digestion (12 bases), was obtained by an analogous procedure using the restriction endonuclease *Rsa* I. The position refers to the actual site on the template DNA where the terminal ATP of the RNA primer is base paired. The size of each *Hae* III fragment refers to the nucleotide length from the terminal ATP of each RNA primer to the nearest *Hae* III cut in the 3′ direction.

All 13 Initiation Sites on ϕX174 Share the Sequence 3′-C-T-G-G-G-5′ or 3′-C-T-G-G-T-5′. Because the sequence complementary to pppA-C-C-C occurs 8 times in ϕX174 and the sequence complementary to pppA-C-C-A occurs 58 times, a gene *4* primer recognition site must consist of more than the sequence complementary to the primer. To determine if there is a more extensive recognition sequence, we determined the sequences of the 5′ ends of all 13 fragments in an analogous manner.

The sequences determined for the 5′ ends of the 13 fragments are shown in Fig. 4. At all of the sites either 3′-T-G-G-G-5′ or 3′-T-G-G-T-5′ is found immediately adjacent to the 5′-terminal nucleotide of the DNA fragment. We conclude, therefore, that both pppA-C-C-C and pppA-C-C-A are used as primers by the gene *4* protein, consistent with published data on the primer sequence (6, 12). In addition, there is invariably a deoxycytidine residue on the 3′ side of the tetranucleotide sequence complementary to the primer sequence. Of the 8 3′-T-G-G-G-5′ sites on ϕX174, only 3 have the sequence 3′-C-T-G-G-G-5′; of the 58 3′-T-G-G-T-5′ sites, only 10 have the sequence 3′-C-T-G-G-T-5′. These 13 sites are all used as initiation sites; no other predominant sites are observed. Thus whenever the gene *4* protein finds the sequence 3′-C-T-G-G-G-5′ or 3′-C-T-G-G-T-5′ on the ϕX174 DNA template, it synthesizes the complementary pppA-C-C-C or pppA-C-C-A primer.

In addition to the 13 predominant initiation sites observed in Fig. 2, longer exposure of the autoradiograph reveals a background of faint bands corresponding to minor initiation sites (not shown). These sites (see *Discussion*) correspond to the more general gene *4* protein recognition sequences 3′-C-T-G-G-N-5′ and 3′-C-T-G-T-N-5′, of which there are 30 on ϕX174 DNA.

Gene *4* Protein Translocates Along Single-Stranded DNA in a Unidirectional 5′-to-3′ Direction. The variation in intensity of the different end-labeled fragments shown in Fig. 2, lane B, is a consequence of differences in utilization of potential primer



Number of bases to next site in 5' direction	1014	24	141
Predicted ratio of utilization	42	1	6
Actual ratio of utilization	26	1	7

FIG. 5. Spatial arrangement of three primer sites and their relative frequency of utilization. Predicted ratios of utilization are based on the model for unidirectional 5'-to-3' translocation by the gene *4* protein (see text). Actual ratios of utilization are based on the direct comparison of radioactivity in the initiation fragments shown in Fig. 2, lane B. In both the predicted and actual ratios, site 4688 has been normalized to 1.

sites. We propose that the gene *4* protein binds to any single-stranded DNA sequence and then translocates in a unidirectional 5'-to-3' direction with regard to the DNA strand until it finds a primer site. Such a mechanism implies that the relative utilization of a primer site is dependent upon the length of DNA between it and the next primer site in the 5' direction. Consider, for example, a 1200-base region of φX174 DNA containing four gene *4* protein recognition sites, separated by 1014, 24, and 149 bases (Fig. 5). If the gene *4* protein bound randomly to this DNA, and then migrated 5' to 3' until it found a primer site, then the relative utilization of the three rightmost sites should be the same as the relative distance to the next primer site in the 5' direction, or 42:1:6. The ratio found was 26:1:7.

Fig. 6 compares the relative utilization of all 13 recognition sites on φX174 DNA. The frequency of usage is plotted against the number of bases to the next primer site in the 5' direction. There is clearly a strong correlation between the observed utilization of a given site and the length of the DNA from that site to the next one located in the 5' direction.

On the basis of the above model, the utilization of primer sites should vary linearly with the number of bases to the next primer sites in the 5' direction. The nonlinearity observed (Fig. 6) would result if the gene *4* protein occasionally fell off the DNA template before priming or if it skipped over potential primer sites with a given frequency.

DISCUSSION

Evidence obtained both *in vivo* and *in vitro* shows that the gene *4* protein is responsible for initiating lagging-strand synthesis by catalyzing the synthesis of RNA primers, which are then extended by T7 DNA polymerase (6, 10–12, 20). Romano and Richardson (12) and Scherzinger *et al.* (6) have shown that the primers synthesized *in vitro* are predominantly tetranucleotides having the sequences pppA-C-C-C and pppA-C-C-A. This result is in agreement with the results of Okazaki *et al.* (21, 22), who have shown that the RNA primers made *in vivo* have the sequence pppA-C$(N)_{2-3}$, in which N is mainly A and C, and the chain length is predominantly four ribonucleotides.

In this paper we have shown that primer synthesis by the gene *4* protein is template dependent. The major recognition sequences for the gene *4* protein are 3'-C-T-G-G-G-5' and 3'-C-T-G-G-T-5', with the resultant synthesis of a pppA-C-C-C or pppA-C-C-A primer. In addition to these predominant sites, the gene *4* protein is also capable of synthesizing primers at additional sites, albeit at a much lower efficiency (less than 10%). Analysis of these minor initiation sites (to be published) shows that the more general gene *4* protein recognition sequence is 3'-C-T-G-G-N-5' or 3'-C-T-G-T-N-5'. Thus it appears that the gene *4* protein recognizes a specific core sequence, 3'-C-T-G-5', while tolerating considerable heterogeneity in the next two nucleotides. The site specificity described here is not affected by the presence of T7 or *E. coli* DNA binding proteins, or by the

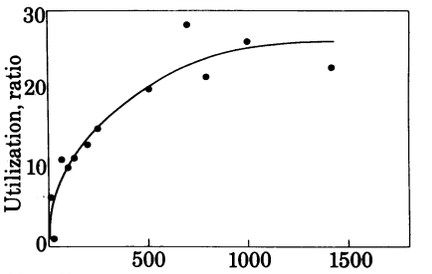


FIG. 6. Utilization of the φX174 primer sites as a function of the distance to the next primer site in the 5' direction. The ordinate represents observed utilizations, expressed as a ratio to the utilization observed for site 4664 (which has been normalized to 1). Data shown represent the average of five different preparations of [γ-^{32}P]ATP-labeled RNA primers digested with either *Hae* III, as shown in Fig. 2, lane B, or the restriction endonuclease *Rsa* I (not shown). Radioactivity was measured either by solubilization of gel slices and scintillation counting or by densitometer tracings of the autoradiographs.

DNA template (unpublished results). Recent *in vivo* studies of Okazaki *et al.* (22) suggest a recognition sequence identical to that found in our *in vitro* studies. Mapping five initiation sites on newly replicated T7 DNA to a region of T7 DNA of known sequence, they infer the recognition sequence for RNA primer formation to be 3'-C-T-G-G-N-5' or 3'-C-T-G-T-N-5'.

An essential feature of the gene *4* protein recognition sequence is the deoxycytidine residue located distally at the 3' end, a nucleotide that is not transcribed into the RNA primer. The presence of this additional nucleotide in the recognition sequence means that, on the average, only a quarter of the sequences complementary to the primer sequence will be recognized by the gene *4* protein. The heterogeneity observed in the sequence of the RNA primer (6, 12, 21) can be accounted for by the recognition sequence. For example, Scherzinger *et al.* (6), using φX174 DNA as a template, found the sequence pppA-C-C-A to predominate over the sequence pppA-C-C-C. Such a bias is expected because there are 10 3'-C-T-G-G-T-5' sequences and only 3 3'-C-T-G-G-G-5'. All 13 of these gene *4* protein recognition sites are used, without preference for 3'-C-T-G-G-T-5' over 3'-C-T-G-G-G-5'.

Heterogeneity has also been observed in the length of the T7 RNA primers (10, 12, 21): they are predominantly tetranucleotides, with a minor proportion of pentanucleotides. Such heterogeneity was also observed in the studies described here. Although the primers synthesized at some sites are exclusively tetranucleotides as judged by the unambiguous sequencing pattern, others, such as that at position 1585, give rise to a second, fainter, sequencing pattern indicative of a mixture of tetranucleotide (90%) and pentanucleotide (10%) primers. Thus the heterogeneity in length of the RNA primer reflects variable length primers synthesized at given sites.

Of the 13 gene *4* protein recognition sequences on φX174 DNA, there is wide variability in their utilization. The relative utilization of a given site appears to depend upon the number of nucleotides that occur between that and the adjacent site in the 5' direction. This phenomenon is most readily explained by a random binding of the gene *4* protein to single-stranded DNA, with subsequent translocation in a 5'-to-3' direction until a primer site is found. Such polar movement has been suggested for other ATP-dependent single-stranded DNA binding proteins. Evidence for unidirectional 5'-to-3' migration on single-stranded DNA exists for the bacteriophage T4 DNA helicase (23) and *E. coli* helicases I, II (23), and III (24), and a similar mechanism has been postulated for the *dnaB* protein (25). The *E. coli rep* protein moves in the opposite, 3'-to-5', direction on single-stranded DNA (24).

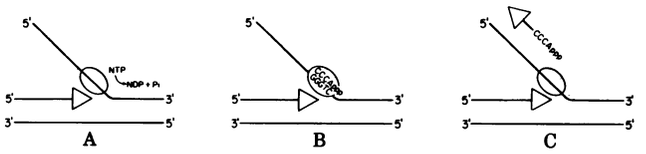


FIG. 7. Model of gene 4 protein activities at a replication fork. (*A*) Gene 4 protein (represented by the ellipse) moves 5′ to 3′ on the displaced strand, stimulating T7 DNA polymerase (represented by the triangle) by facilitating unwinding of DNA at the replication fork. (*B*) At the recognition site, gene 4 protein synthesizes RNA primer. (*C*) RNA primer is extended by T7 DNA polymerase to initiate lagging-strand synthesis.

A model correlating the diverse enzymatic activities of the gene 4 protein with the different functions required at a replication fork is shown in Fig. 7. In this model the gene 4 protein is bound to the single-stranded lagging strand at the fork in juxtaposition to the T7 DNA polymerase, which is bound to the leading strand. The gene 4 protein, either alone or in concerted action with the DNA polymerase and DNA synthesis, translocates in a 5′-to-3′ direction on the lagging strand. The translocation would presumably require the hydrolysis of NTPs, facilitating the unwinding of the helix and hence the movement of the fork. When the replication fork arrives at a gene 4 protein recognition site on the displaced strand, the gene 4 protein binds to this sequence and synthesizes a complementary tetraribonucleotide primer, and lagging-strand synthesis is initiated in the opposite direction.

Primer synthesis by the gene 4 protein may well play a pivotal role in the initiation of replication at the primary origin of the T7 chromosome. We have recently shown that the primary origin of T7 DNA replication is located within a 100 base-pair region 15% of the distance from the left end of the T7 DNA molecule (26, 27). This region contains an intergenic 70-base-pair A+T-rich sequence within which is a single gene 4 protein recognition site (3′-C-T-G-G-G-5′). If this A+T-rich region were melted out by transcription through the region by T7 RNA polymerase starting at the two tandem promoters immediately to the left, then the gene 4 protein recognition site would be exposed. Primer synthesis at this site would initiate DNA synthesis in a rightward direction. Another gene 4 protein molecule could then bind to the lagging strand and translocate in a 5′-to-3′ direction until it reached the fork, at which point the events depicted in Fig. 7 would be initiated.

On a random DNA sequence, predominant gene 4 protein recognition sites occur with an average frequency of once per 500 bases. However, the length of Okazaki fragments in T7 range between 1000 and 6000 bases (12, 22). This discrepancy could be explained by the presence of fewer gene 4 protein recognition sites on T7 DNA than would be predicted by a random distribution of nucleotides. Alternatively, the gene 4 protein may synthesize RNA primers infrequently at a replication fork, skipping over potential sites in most instances. This clearly is the case for the minor class of gene 4 protein recognition sites on ϕX174 DNA, as discussed in this communication.

At present the mechanism of RNA primer formation and discontinuous DNA replication is best characterized for the phage T7 system (22). Some phage, (e.g., M13) use the *E. coli* RNA polymerase for priming of DNA synthesis. Other phage (e.g., G4), as well as *E. coli*, use the *E. coli dnaG* protein; at recognition sequences that possess extensive hairpin structures this primase synthesizes a mixed RNA-DNA primer of heterogenous length. Phage T4, like T7, synthesizes its own priming proteins (genes *41* and *61*) (28, 29). RNA primers found at the 5′ ends of T4 DNA are very similar in sequence to T7 primers: they are mainly pentanucleotides, with the sequence pppA-C(N)$_3$ (21, 28–30). Recognition sequences have not yet been determined for the T4 primase, although evidence does suggest that priming is site specific (28, 30).

In addition to playing an essential role in T7 DNA replication, the site-specific priming activity of the gene 4 protein should provide the basis for novel methods of *in vitro* analysis of DNA. The gene 4 protein could be used to generate unqiue fragments for DNA sequence determination and as a tool to generate site-specific mutations at gene 4 protein recognition sequences.

This investigation was supported by U.S. Public Health Service Grant AI06045 and American Cancer Society Grant NP-1J.

1. Richardson, C. C., Romano, L. J., Kolodner, R., LeClerc, J. E., Tamanoi, F., Engler, M. J., Dean, F. B. & Richardson, D. S. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 427–440.
2. Modrich, P. & Richardson, C. C. (1975) *J. Biol. Chem.* **250**, 5508–5514.
3. Modrich, P. & Richardson, C. C. (1975) *J. Biol. Chem.* **250**, 5515–5522.
4. Mark, D. F. & Richardson, C. C. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 780–784.
5. Kolodner, R., Masamune, Y., LeClerc, J. E. & Richardson, C. C. (1978) *J. Biol. Chem.* **253**, 566–573.
6. Scherzinger, E., Lanka, E., Morelli, G., Seiffert, D. & Yuki, A. (1977) *Eur. J. Biochem.* **72**, 543–558.
7. Hinkle, D. C. & Richardson, C. C. (1975) *J. Biol. Chem.* **250**, 5523–5529.
8. Scherzinger, E. & Klotz, G. (1975) *Mol. Gen. Genet.* **141**, 233–249.
9. Kolodner, R. & Richardson, C. C. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 1525–1529.
10. Scherzinger, E., Lanka, E. & Hillenbrand, G. (1977) *Nucleic Acids Res.* **4**, 4151–4163.
11. Romano, L. J. & Richardson, C. C. (1979) *J. Biol. Chem.* **254**, 10476–10482.
12. Romano, L. J. & Richardson, C. C. (1979) *J. Biol. Chem.* **254**, 10483–10489.
13. Deutscher, M. P. (1973) *Prog. Nucleic Acid Res. Mol. Biol.* **13**, 51–92.
14. Hutchison, C. A., III & Sinsheimer, R. L. (1966) *J. Mol. Biol.* **18**, 429–447.
15. Adler, S. & Modrich, P. (1979) *J. Biol. Chem.* **254**, 11605–11614.
16. Weiss, B., Live, T. R. & Richardson, C. C. (1968) *J. Biol. Chem.* **243**, 4530–4542.
17. Okazaki, R., Hirose, S., Okazaki, T., Ogawa, T. & Kurosawa, Y. (1975) *Biochem. Biophys. Res. Commun.* **62**, 1018–1024.
18. Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchison, C. A., III, Slocombe, P. M. & Smith, M. (1978) *J. Mol. Biol.* **125**, 225–246.
19. Maxam, A. M. & Gilbert, W. (1980) *Methods. Enzymol.* **65**, 499–560.
20. Wolfson, J. & Dressler, D. (1979) *J. Biol. Chem.* **254**, 10490–10495.
21. Okazaki, T., Kurosawa, Y., Ogawa, T., Seki, T., Shinozaki, K., Hirose, S., Fujiyama, A., Kohara, Y., Machida, Y., Tamanoi, F. & Hozumi, T. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 203–219.
22. Ogawa, T. & Okazaki, T. (1980) *Annu. Rev. Biochem.* **49**, 421–457.
23. Kuhn, B., Abdel-Monem, M., Krell, H. & Hoffmann-Berling, H. (1979) *J. Biol. Chem.* **254**, 11343–11350.
24. Yarranton, G. T., Das, R. H. & Gefter, M. L. (1979) *J. Biol. Chem.* **254**, 12002–12006.
25. McMacken, R., Ueda, K. & Kornberg, A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4190–4194.
26. Tamanoi, F., Saito, H. & Richardson, C. C. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2656–2660.
27. Saito, H., Tabor, S., Tamanoi, F. & Richardson, C. C. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3917–3921.
28. Liu, C. C., Burke, R. L., Hibner, U., Barry, J. & Alberts, B. (1979). *Cold Spring Harbor Symp. Quant. Biol.* **43**, 469–487.
29. Silver, L. L. & Nossal, N. G. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 489–494.
30. Nossal, N. G. (1980) *J. Biol. Chem.* **255**, 2176–2182.